DOCUMENT RESUME

ED 186 711 CE 025 299

AUTHOR Baker, Eva L.
TITLE New Directions in Evaluation Research: Implications for Vocational Education. Occasional Paper No. 55.
INSTITUTION Ohio State Univ., Columbus. National Center for Research in Vocational Education.
PUB DATE Nov 79
NOTE 18p.; Paper presented at The National Center for Research in Vocational Education (Columbus, OH, 1979).
AVAILABLE FROM National Center Publications, The National Center for Research in Vocational Education, The Ohio State University, 1960 Kenny Rd., Columbus, OH 43210 ($1.90)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Assessment; *Educational Research; Educational Researchers; *Evaluation Methods; Evaluation Needs; *Government Role; *Measurement Objectives; Politics; Program Evaluation; Research Methodology; Speeches; *Vocational Education

ABSTRACT

In the early 1970s educational evaluation's fusion of respect for rationality, power to implement rational procedure, and an altruistic objective (improved instruction) appealed to educational researchers. Dealing essentially with closed systems, it measured program success by student performance on measurement instruments. Critics exclaimed that program developers and evaluators took away personal decision rights from students and that outcome measures were incomplete and inaccurate. Changes occured due to reanalysis showing the futility of earlier research, operating focus on procedures (not outcomes), open student participation, and loss of stability required for longitudinal study. In present-day open system evaluation settings, emphasis is on multiple objectives, and selection of what is to be evaluated is left open. No longer committed to methodology and provision of clear information, evaluators use case studies and looser, more interactive designs. Their reactions to political demands are to use needs assessments, develop systematic procedures, and write more reports. Problems between politics and evaluation include these: (1) the technical disagreement that any evaluation is subject to erodes its credibility with its contracting agency, (2) evaluation results can be used to discredit politicians whose claims outstrip their programs, and (3) success of political efforts gives politicians a sense of personal power. (YLB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED186711

Occasional Paper No. 55

NEW DIRECTIONS IN EVALUATION RESEARCH: IMPLICATIONS FOR VOCATIONAL EDUCATION

by

Eva L. Baker

Director of the Center for the
Study of Evaluation
Los Angeles, California

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Joel Magisos

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

The National Center for Research in Vocational Education
The Ohio State University
Columbus, Ohio 43210

November 1979

CE 023 297

PREFACE

We are pleased to share with you a presentation entitled, "New Directions in Evaluation Research: Implications for Vocational Education," by Eva L. Baker, professor of education at UCLA and director of the Center for the Study of Evaluation in Los Angeles, California.

In her speech, Dr. Baker outlined a brief history of evaluation efforts from the early '70s to the present and shared with us some of her personal experiences as an educational researcher and evaluator. She pointed out that the changes in the political climate of the country have had far-reaching implications for education in general and especially for evaluation methodology and rationale.

Dr. Baker received her A.B., M.A., and Ed.D. degrees from the University of California at Los Angeles. She brings to her present position as director of the Center for the Study of Evaluation an extensive background in university teaching, curriculum design, staff development activity, and educational research. She serves on the editorial boards of several prestigious journals and is affiliated with numerous professional organizations in the field of educational development and research.

We are grateful to Dr. Baker for sharing with us her experiences and insights which proved timely and thought-provoking. We are proud to present her lecture entitled, "New Directions in Evaluation Research: Implications for Vocational Education."

Robert E. Taylor
Executive Director
National Center for Research
in Vocational Education

NEW DIRECTIONS IN EVALUATION RESEARCH: IMPLICATIONS FOR VOCATIONAL EDUCATION

I would like to talk about the values underlying educational evaluation. I don't want to talk about them theoretically, but rather from what I think is a very clear-cut point of view. I believe that values are inherent in what people *do,* not just in what they say. If we study the *actions* we're involved in we can find out something about the values of evaluation. Clearly, evaluation activity is made up of roles that different people play, sometimes at different times. I think the tasks of evaluation vary significantly depending on the kind of program we happen to be studying. So I think the idea of a full and perfect generalizable methodology is unfair if not impossible.

If you look at the work that evaluation people have done over the last fifteen years, you can see qualitative changes in what was expected, what was promised, and what was delivered. It's very enlightening to go back to some earlier evaluation reports which were regarded as satisfactory documents and compare them with the kinds of things that are coming out now. To put all this in context, I think it's important to understand a little about the history of evaluation – principally evaluation as seen by researchers or research professionals, since most of you fit that description.

When evaluation activity became highly visible (at least ten years ago), it appealed to values that were held by many educators. The skills in research that we had acquired were modeled more from the science end of education rather than the art end, and we thought those skills could be applied in a fashion that would lead us to the improvement of opportunities and desired outcomes for students. So it seemed possible to fuse several elements – first, a respect for rationality, which as researchers we presumably had; second, the power to implement rational procedure, which as researchers we probably didn't have, at least in a political sense; third, an altruistic objective, which we saw as improved instructional life for students; last (and not incidentally for some), a supplementary source of livelihood, because the short, happy life of well-supported research activity was over.

It seems to me that the short life of research activity – of the cooperative research program – was important in two respects. First of all, the brevity of that period bred intensity and commitment. People developed their research skills. There were research training programs in the middle sixties that were funded out of Title IV, the same enabling legislation that generated the labs and center program. Much research was being done. As the research opportunities lessened, the researchers became evaluators, an interesting job change. People assumed they could apply their research rationality to evaluation problems. This view implied optimism. It was assumed that evaluation shared the basic precepts of a good science – that evaluation was independent the way research is supposed to be independent; that evaluation was orderly; that expertise was required for its conduct; and that, by virtue of training, some people ought to be better at it than others. Supporting beliefs honored the value of measurement, attractability, discovery of causal relationships, suspension of disbelief for questionable human data sources and, of course, the idea of design and control. All these research ideas were transferred to the evaluation framework.

Evaluation also proceeded in an atmosphere of mutual support. Optimism was based not only on the idea that we could identify a treatment or determine whether or not a program was any good, but evaluation was also to contribute to the productivity of such programs. In general, these beliefs were shared by program developers and managers in evaluation. Government people, those in fact responsible for contract management, were generally less well trained and were usually inexperienced as well, so they were moderately comfortable with a well-argued plan of action that promised substantial benefits.

From an instructional point of view, sometimes these premises were valid. We would try out some of the theories – for example, the work that Lumsdaine and Glaser did on deciding how instructional programs ought to be developed and tested. Pre-tests were developed, instructional treatments were implemented, data were collected, and revision cycles recurred. We had the idea of "Social Darwinism" – that in some sense, we were getting better and better each time as we collected data. In fact, there was some research evidence to support this. This is the kind of model through which I got into evaluation. It's a time-series design; we collected data and tried to make things better.

Obviously those of you who have some background in systems analysis recognize that the kind of work we were doing dealt essentially with closed systems where there wasn't a whole lot of uncertainty. Program developers at that time, and I'm now talking about 1970-1971, were still in relatively good control of the population of learners that they were dealing with. They could exclude or include people by pre-test, and they were able to control pretty much what happened instructionally in the treatment. As I said earlier, sometimes the desired learning did occur; and at that point we made inferences about how good a program was by looking at student performance on measurement instruments. Many of you can recall the evangelical fervor with which some people pushed instructional objectives. (I was probably in that set.) The concept of learning through instructional objectives was based on the idea that student performance ought to follow from the instruction that is presented to them. Very often, however, we didn't ask anything about the students except the perfunctory, "How did you like the program?" We weren't centrally interested in the long-range effects of the program or in any serious attitude change. At most, we were interested in instrumental information that would allow us to make the program better next time.

There were, however, "voices from the wilderness" (and many of us thought they should remain there) that objected to the overall strategy. (At this time I was at the Southwest Regional Lab working on the development of their reading program, and I was into the cycle in a big way.) Critics of what we were doing identified the "top down" nature of instructional development and noted that many personal decision rights were preempted by the developer and the evaluator and taken away from the students. Claims were also made that our outcomes measures were, in any case, incomplete, and probably inaccurate as well. Other considerations were voiced but summarily dismissed in large part, I would guess, for the wrong reasons. One reason these criticisms seemed so easy to dismiss is that similar kinds of criticisms were coming from people who were self-avowed "protectors of humanism." Some of you may recall that the neo-humanistic movement caused a great deal of controversy. The people involved in this movement were anti-technology, anti-schooling, and frequently associated with encounter groups and Esalen-type experiences. They were "typical Californians," I might say – certainly a group that's easy to discredit. Another reason for overlooking the criticism, which I must admit with some embarrassment, was my own personal reason: the criticism seemed to attack my sense of personal accomplishment. I couldn't help but think, "By damn, I know I'm doing something. Why are they putting me down?" In this closed system kind of evaluation, we thought that instructional strategies were, very loosely, perfectable. We would also try to improve our procedures, and we thought these, too, were improvable.

The major conceptual distinction between evaluation and our former line of work, research, was in the treatment of generalizations. At the heart of most scientific research was the effort to find new knowledge and generalize it; at the heart of most instructional development experiences was the effort to find information about a particular program and generalize it to other similar population groups. It seems to me that some of us searched in vain for regularities which might allow us to consolidate and improve our methods, and I did a lot of writing about that at one point. But such generalizations about procedures would be frosting on the cake; I might add that it would be pink fondant roses if we found R&D procedures of general use. But there was a sort of reluctance or inability at that point to formulate those ideas in a way that could be shared.

Now to show you something of the transition between what evaluation was like up to about 1973 and what evaluation is like now, I am going to describe a little of it, six years later. Whether the change came about because we have a different frame of mind or because there are broad, socially-inspired shifts that have occurred in the meantime (including the lack of faith which President Carter's speech identified), nonetheless, there has been a change. I think part of this was supported by the re-analysis of the studies of schooling done by Coleman and others which tended to point out that most of what we did was futile at the margin, and there weren't any changes taking place that couldn't be explained better by demographic information. We had a spate of NSD ("no significant difference") research, even though in our own minds we would imagine that there were very large treatment differences. The level of resources had changed dramatically, but we just weren't finding any of those changes reflected in measurement. At the same time, federal research and development activity was severely constrained, partly because of overall shifts in the government and partly, I think, because of the ineptness of some of the appeals that went on at that time. I would urge you to read the book by Sproull, Weiner, and Wolf called, *Organizing an Anarchy*, which is pertinent to this point.

At any rate, there were also different views on resource allocation at the legislative level, and ideas such as zero-based budgeting required presumably tougher tests to be applied to programs than we had before. At the same time, concern for educational equity, led by the courts, generated a set of programs which might in fact be legitimate just by their very existence rather than by their effects. The issue was equity of opportunity; not equity of outcome. Programs for bilingual children, for example, could survive perhaps very negative evaluations because there was clearly something that needed to be done by the government in that sector. Educational programs became, in some situations, vehicles through which to reallocate resources rather than real treatments. This further shifted the operating focus from outcomes to procedures. The educational systems that were now addressed were open instead of closed with regard to the nature of the programs undertaken. Many more local options were provided. The nature of the participation of students was more open, since it was very difficult to restrict or exclude students on the basis of not meeting the entry level criteria set up by the program. Students were willing or unwilling participants depending on where they were in the transiency or mobility bands around the school. The final point is that the basic stability required for the identification and the evaluation of educational activity just wasn't there, especially in a longitudinal framework. If a transiency rate is 20 percent annually, over three years the turnover is substantial. It's very difficult to do a longitudinal study now and have the idea that one is dealing at all with the same cohort.

So apparently a new set of values has become adopted by people who do evaluation in more open system evaluation settings, and I think there are some very significant differences which have implications for how we act. One difference is that there is a new emphasis on pluralism, on diversity, and on multiple objectives. Also, the selection of what will be incorporated into the program is left wide open and very often subject to local preference, since it is argued that the

findings of educational research have failed to give priority to alternative courses of action anyway. The resultant mix of activity in schools can hardly be called a program at all. (I may exclude vocational education programs if you have better information. I'm speaking primarily about what I know goes on in the general education programs in the public schools.) The notions of treatment and the attendant links with causality are concepts that we can hardly deal with anymore. Refinement of programs at the level of precision which characterized many of the curriculum development efforts of the early labs and centers is really beyond comprehension in most of these programs now, especially in urban settings. This is difficult for some to comprehend because much of our language has stayed the same. We still talk about the programs as if they were the same entities we had five or six years ago that you could hold onto in some ways, describe, and manipulate.

At the same time, evaluation roles have changed, too. Evaluators in the sixties, whether they were looking at instructional units or broader-based policy efforts, were very much committed to methodology and to the provision of clear information – to the use of a science base. The initial response of evaluators to these changing requirements of open systems and more diffuse views of education was, as you might imagine, to search for better measures and better techniques – that is, to look first to methodology as a way of solving their problem. So some of us began looking for better ways of aggregating information, refining our designs, refining alternative methodologies, and conducting comparative studies. There were also other tactics taken during this transition. For instance, the question was raised about the utility of different types of data and trade-offs for data reliability. We could achieve a sense of consistency by viewing data across sites, so we made greater use of case study versus survey kinds of research. Preferences developed among some evaluators for looser, more interactive designs, and what I would call "soft" data sets; and sides were chosen – "hard" opposed to "soft" – although maybe these alignments aren't necessary and are probably dysfunctional. Radical approaches – ideas Bob Stake talked about in 1972, or Bob Rippey outlined in 1973, and Egon Guba proposed in 1978 – cast the evaluator more as a responsive inquirer than as a provider of purely objective views. Critics claimed this new responsiveness was only labeling and legitimizing what was the case anyhow. They argued that we were already biased, so why didn't we just name our biases? That is, they saw the evaluator entering with screens through which the data and perceptions pass. Somehow we assumed that this screening would "randomize out" through use of a great number of case studies. The participant evaluator role was also conceived as a foil to the role of summative evaluator with the latter, summative evaluation, being assumed to adhere strictly to comparison and choice among program options.

Other questions were raised, of course, about the objectivity of evaluation methodology. Henry Aaron, from the Department of Health, Education, and Welfare, pointed out in one of his speeches that evaluation methodology is inherently conservative. It gives a tough test of differences because that's the way the statistical paradigm is structured. In other words, the structure of evaluation methodology is prone to give a finding of "no significant differences" when, in fact, there may be differences which are ignored or overlooked. So there is and will continue to be considerable debate about the best roles evaluators should take, and the type of data most useful, and so on. It seems to me that these changes were gradual. I'm making them seem more dramatic because I'm looking at them in retrospect. These changes also seem to be characteristic of changes in other disciplines.

At this point I would like to discuss the evaluation community's reaction to the specter of politics, a factor which grew in our awareness each year. By politics I'm talking not only about major legislation and national politics, but politics at every level. It seems to me that when the discussion and execution of evaluation entered the area of politics, the first set of responses we made as evaluators was wholly predictable. We thought we should implement old and important

values in an effort to control the situation since, as evaluators and researchers, we were interested in control. We responded to the politics of the situation somewhat like a paramecium that ingests by envelopment – we tried to confront politics by surrounding and absorbing it. We thought we could gain control of what we saw as a political incursion into our area. Our first thought was, "Why are they messing with us?" We also thought we could manage what we saw as the "irrational" side. What we wanted to do was to bring that whole set of experiences within the boundaries of "mainline" educational R&D.

Let me give you one or two examples: To deal with the conflicting goals, a political problem which was being loudly articulated in the very early 1970s by various constituencies with interests in school programs, evaluators borrowed the needs assessment idea from the community development people in sociology. So the evaluation profession's response to politics at that time could be characterized by such statements as "We can solve that problem by giving everybody an opportunity to say what they want to say in a needs assessment, and that will keep it nice and neat." We developed systematic procedures. The idea was that we were going to have an expression of pluralistic views in an attempt to control the "politicizing" of evaluation findings, which meant, in my view, that most evaluation work was being used only as a device to argue for or against favorite programs. Evaluators developed and promoted adversary or advocacy forms of evaluation where pros and cons were pitted against each other. I'm sure you know the work of Bob Wolf, Marilyn Kourilsky, and a number of other people who have tried this kind of technique. In the market research area, we borrowed the idea that we had to write reports – different reports for different constituencies – because it was important to keep building our connections with those constituencies. So we started generating more paper.

What those illustrations show is that people in evaluation tried to transform existing societal and political reality into procedures that could be controlled, and we have perpetuated some anomalies on the body of education by doing that. We have community advisory committees providing input, and we do needs assessments – some school districts do thirteen of them a year, one for each special program they are involved in. In California, we have school site councils which are supposed to provide continuing input into assessment of local programs. We allow the articulation of reams of precise objectives. When I was here six years ago, you had shelves of vocational education objectives; we at the California R&D center had our 578,000 reading objectives, and the Wisconsin R&D centers had theirs. Writing objectives was our way of trying to account for pluralistic views. We reasoned that people would think of vocational education, or reading, or some other area as consisting of alternative views, and the only way we could articulate those views was by writing objectives. Now let's observe an individual case. Suppose you have an evaluator who is trained in educational research. That evaluator confronts a program which values diversity over performance, which values the distribution of program over treatment, and which values activity over outcome. What this evaluator tries to do is to make some sense out of all that, and he or she finds, from the point of view of the educational researcher, that this is an impossible task.

I finally discovered a book by Aaron Wildavsky, a political scientist, who makes the point that politics and planning (planning is his word for any systematic intellectual activity) are equally rational. It isn't that the researchers are rational and the politicians are irrational, but that the norms of planning or evaluation contrast with the norms of politics. He points out that the norms of planning (i.e., evaluation) are methodological norms and exist without content. And the norms of politics have *only* content; that is, what you try to do in a political situation is to agree upon something in the content area. We have, as evaluators, mounted some very interesting alternative responses to the political view of evaluation. Some people in the field argue that the uses of evaluation are primarily political; and we should, therefore, put politics first and use evaluation as

persuasion. They say that our primary effort should be to convince people of the value of certain educational programs if, indeed, we believe that those programs have value. Other people have gone into other lines of work, back to the luxury of small-scale laboratory experimentation where a referee is required only at the journal level. Many have persisted, and I feel I belong in this group; but I may be moving into the first one soon.

Wildavsky makes some claims which may or may not be true, but he says that the social policy researchers did not have to experience the same shock of context and change which people in education experienced. Education has always had a political side, but the researchers have generally been insulated from it. Only recently has this changed. The social policy people do not normally assume to control outcomes (with the exception of Henry Kissinger). Policy people are different from educators because they have not very often had that heady, indelible experience of seeing performance levels for kids change from 60 percent to 90 percent because of something they did. There are some differences in our background, too. Policy analysts are trained in the notion that politics is a reality from start to finish and not something transmuted into life from the front pages of the morning newspaper. Some of the policy analysts even appear to think that politics is fun, not an incursion, and that is a concept which is hard for me to cope with. In contrast to the educational researcher's response, their point is not to get control of politics, but to let it happen. Their goal is *not* to quiet it down, not to strip the work of biases, not to make goals and findings crystal clear and able to stand on their own merits – all the things educational researchers talk about doing. Instead, policy analysts acknowledge and embrace political reality. They make decisions in terms of how they will "play to appropriate audiences." Ambiguity, which they see as allowing both personal and satisfying interpretation, is not always a fault; sometimes they push it. Goals are multiple, outcomes are blurred, and people can feel that their own priorities are taken into account. It seems to me that these policy analysts have taken a particular point of view – a view which you may have some moral problems with, but a view, nonetheless – of how one merges scientific or rational planning with the political reality.

It's interesting to me that there are methodologists in our field who have been able to do this themselves. One of those people is Cronbach at Stanford. In his most recent monographs he's been writing mostly about the political problems as well as the methodological differences that evaluators have to deal with. He seems to be a person who can transcend some of those issues. There are a number of people who can't do that, and these people have done some really strange things. For example, there are those in evaluation who make political connections. Such a person might form a connection with a certain politician and the two together form a kind of hybrid person. One person can't deal with both politics and rationality, so the two form a team. They're like Siamese twins. Other people have decided that they will "sit at the foot of" the contracting agency and try to provide information.

We have an interesting problem with our contracting agencies, and I'm dealing not only with NIE, although they're my favorite agency. First, at the agency level, there is an erosion of belief in the expertise of the evaluator to make unchallengeable technical decisions. The "hired gun" strategy which Mike Patton talked about in his book used to be a device which demonstrated the proposition that evaluation people, like education professionals in general, love to disagree on both major and minor points. The fact that we do this so often and so publicly makes the credibility of anything we do subject to attack, and everyone knows that any study can be ripped apart depending on the point of view of the critic. Any evaluation is subject to some technical disagreement, and this erodes credibility. Credibility is one of the most important coins that a person has in a political context, so that puts us one down. Second, there is a realization that evaluation offers a terrific means for attacking individuals who seem to be above or insulated from more typical approaches

at discredit. Let me expand on this a little. We have society to thank or blame for the fact that our usual modes of discrediting people aren't working anymore. Politicians used to be discredited on the basis of their marital status, sexual behaviors, substance abuse, or fraud. But these seem not to have the cogency that they had in the past. In fact, a clear apology seems all that's necessary even in the face of indictment. You can't get rid of somebody as easily as you could in the past. Evaluation findings now can be used as a supplement to discredit public officials whose rhetorical claims, as usual, outstrip their programs. So evaluation has become an important political tool.

I'll give you a short example. Jerry Brown, the governor of California, had a pretty good idea about linking the university systems and the public school systems by way of a stationary satellite. It wasn't going to cost that much and there were all kinds of reasons to do it. The problem was that those who opposed the program linked it to what they considered some of his "wilder" ideas. They used this particular satellite program to discredit him and his political ideas in general, and it marked the beginning of the continuing problem he's been having with the state legislature in California. That's an example of how an evaluation activity can really discredit a person. The *intent* here was to discredit. Since Jerry Brown is politically astute, he got by.

These developments have had several results. The first is the belief among our contracting agencies that technical dispute equals arbitrariness, so they came to believe that any method is a good method because all methods are likely to generate complaints. Second, it has become evident that evaluation results can depose power; and third, the successes of these combined political efforts have given the politicians and bureaucrats a sense of personal power. They now "understand" the heretofore arcane procedures of research and evaluation. This is a very scary phenomenon.

Let me give you a couple of ways that you see it exhibited. One is in the quality of the RFPs that you are receiving. If your sentiments are the same as ours, those RFPs are prescriptive to the point of being nonsensical; that is, the sponsoring agencies are now telling us what sample procedures to use and what phases to go through. It used to be that they bought your brain; now it's just your arms and legs. What they really want is research assistants – robots to do their work out in the field and then to provide them with results which they can decide how to "patty cake" into good shape. That is one serious way of looking at the notion that the contract agencies know better than researchers and evaluators how to do educational research, because obviously everything is subject to dispute. Sponsoring agencies think they know everything because of the educators' discussions with them. Here is a second example, which is a really wonderful example in some ways. We were working on a proposal last summer for the state of California – a million dollar contract. About twenty days into the response period, we got an amendment by telegram advising us that the California state legislature had mandated control groups. You have to think about that. Here is a political body voting on whether you should have control groups in a research study. That's scary in some ways if you think about independence, lack of bias, the quality of the kind of intellectual rationality you want to bring to bear on educational research.

What do I expect to happen with all of this? I'll tell you, and most of this is taken from Aaron Wildavsky in a book which I recommend to everybody because besides being informative, it's fun to read. It's called *Speaking Truth to Power*, which I think is a great title. He has a chapter called "Strategic Retreat on Objectives; Learning From Failure in American Public Policy." What he says is that public policy is in a similar state across many social/policy areas, not just education. Our early optimism, in which we were going to change the outcomes significantly and solve all the social problems, just didn't pan out; and everybody is feeling upset about this. But he feels that what is happening now is a translation to concern not with outcomes but with process. The evaluation literature today is greatly concerned with implementation evaluation, and that simply means making sure that the process happened. A type of study illustrating this is Milbrey McLaughlin and Paul Burman's Rand studies on innovation.

Wildavsky sees this as a strategic retreat from objective evaluation. He thinks what's going to happen next is that programs are going to be legislatively formulated exclusively in terms of the services they provide, and the rhetoric about what they're supposed to do is going to be dropped. That has a lot of implications for people in evaluation.

What does all this mean? I think it means that evaluation problems might be easier if we could come to some kind of agreement among the evaluators, the program managers, and the contracting agencies. If we could agree that activity is a legitimate way to assess the issue of innovation in educational settings, then the effect of this realignment of value perspectives may work to permit evaluators and their work once again to hold some trusted status.

QUESTIONS AND ANSWERS

Question: Would you please comment on what you believe will be the direction of educational evaluation over the next ten years?

I'll tell you what I'd like to say. I'm a believer in outcomes. I think that a lot of the critics of quantitative-oriented evaluation studies are absolutely correct, that the basis on which judgments were made and no significant differences were found was mostly attributable to poor dependent measures – that is, bad post-tests. I would take the view that the measurement process, if it's going to be treated seriously in these evaluations and not swamped by this emphasis on service, is going to have to be made a lot more credible.

One way of making the measurement process credible is to make it publicly accessible. Up to this point, certain people have gotten away with the idea that they were arcane elves in Princeton making up instruments that everybody could exclaim over, saying, "Yes, if you say this measures my academic achievement, it does!" If you look at the legislative trends toward public access to tests in Texas, New York, and Massachusetts (these were introduced nationally a couple of times), there's a concern that the public needs to have access to these tests, particularly for a constitutional purpose, a due process purpose. It's only if the students and teachers know what's going to be on a test that the proper services can be legally assured. That's the basis on which the Florida case was argued. So our interest in this is to get hold of the measurement side now, and to do it at the point where most people in school districts think evaluation exists. We learned from a study we did that about 75 percent of the people in evaluation research units in school districts think that evaluation is testing. They see those things as isomorphic. What we would like to do is to have the tests exhibit certain criteria. One of them is public access, and we think the way to do that is through specifications – not through the annual publication of all the tests. Zacharias at MIT has computed the core computer space it would take to publish all of the test items annually. Our main concern is with public access.

Our second concern is with money and conservation of resources. We're doing a study on the NIE grant (at least we hope we're doing a study on the NIE grant) which is a survey of all testing practices in public schools. Our understanding of it now is that tests are being regarded as dysfunctional in the public schools. So now we're thinking in terms of public access, the economy issue, the notion that tests need to relate to instructional programs very directly, and the idea that teachers and kids have to regard tests as meaningful and important activities – that is, as measures that have what we used to call "face validity." This would be some way of grabbing hold of the evaluation process. If that doesn't happen, what we have to start doing is finding measures and indicators that have common sense appeal to people. You see, people in testing sometimes argue for tests on the basis of their validity coefficients, which doesn't make much sense to a lot of people. We have to produce meaningful tests that make sense to people. That's the only way we're going to be able to keep the hook into outcomes that we need. That's over the short term. Over the long term everything will switch back, of course.

Question: What about the role of evaluation in terms of policy?

Well, I think it's absolutely important for us to try to maintain an independent stance in evaluation. What we tried in California was a terrible mistake. We were a state university evaluating a state agency on which the superintendent of that agency was a member of the Board of Regents who also happened to be on the Advisory Board to NIE. We were coming and going. What you end up doing is getting everybody choosing up their favorite team and getting an evaluation together that says what the politicians want it to say.

In a political context, I'd like to speak about summative versus formative evaluation, if I can, in a longer perspective. And from that perspective, I think summative evaluation is a complete waste of time because nothing that's created ever really gets dropped; it just gets transformed or renamed. Summative evaluation is irrelevant except in little tiny program comparisons. But in large-scale Title I type programs, for instance, it's only formative evaluation that counts. It seems to me that the kinds of things that we should focus on are outcomes and processes that can be manipulated, that we can do something about. We should not allow educators who would prefer to be mathematicians or psychologists, those who got into education just because they needed a job, to be the people who decide the important issues. Those people who are conducting evaluations should be forced to attend to outcomes and processes over which the schools and the educational professionals have some input. I think we've been wiped out mostly by data collected about information we have no control over, nor will we ever have any control over it. That's been done in the name of science and in the name of comprehensiveness, which I know is a theme. I would push towards very little evaluation – certainly much less evaluation than we're doing now within each program, but evaluation targeted so that the information could have direct input to somebody else. I wouldn't collect evaluation information for the sake of having a "pretty report." Right now most of evaluation is done for the "pretty report."

Question: What about the long-range evaluation of programs like Project Headstart? You could be expected to evaluate the educational outcomes, but how would you evaluate the non-educational, for instance, the health aspects of such a program?

Well, I probably wouldn't; and I would probably question those data. I have some concern about the correlational value, how those kids were selected and whether they were likely to be healthier, and so on. In general, I think that there's too much emphasis on evaluation right now. I think we should spend a lot less money on evaluation and a lot less money on testing.

Question: Why did you choose to quote Aaron Wildavsky so extensively in relation to policy analysis?

Why Aaron Wildavsky and not someone else? Essentially for reasons that probably have more to do with happenstance and not for the best reasons of all. Aaron came at my invitation to the R&D center as a visiting scholar. He conducted a seminar for members of the state legislature and their staff people on what research had to say to policy. And why did I ask him to come as a visiting scholar? Because I asked the legislators who in the California university system they would most like to have come to speak to them. They requested Aaron Wildavsky. So first of all, I had a personal connection. Secondly, it seems to me that he talks in very practical terms about evaluation and policy. I regard myself as an activist. It just seems to me that Wildavsky had an activist's approach. I know it's incremental, and I know what he advocates is different from other social policy approaches. But that was the basis – no lengthy library search.

Question: No one seems to agree about outcomes. Instead, they seem to agree on the *process* or activity. How, then, can you focus on outcomes?

I suppose that my background shows, which is research on learning and instruction. I accept much of what you say. I've had quite a bit of experience with teachers who are dealing with some really basic areas like reading, writing, and math. We have a research project at the R&D center on writing assessment. It's remarkable how people in English departments can agree on what the criteria are for good writing. So I think it's possible to find places where people agree. What I think we should do is exploit those now. It may be that when we start getting into some general, conceptual matter, where we attempt to define it and assess a person's understanding of it in an X or Y manner, we'll end up with fragmentation. I am not an advocate of outcomes or objectives for every area of instruction or every kind of instructional program. In some cases, I think that the best we can do is to show people that they have greater flexibility to react to a variety of indicators or outcomes than they had before and not even worry about what those are. But for the areas where we still haven't demonstrated to the satisfaction of the electorate and the public that we know something about them, like reading and writing instruction, I think we should get hold of those now and not just let them go.

The interactions, if you want to think statistically for a minute, are really impressive. As a measure of instruction, let's take writing for an example. I don't know if this is the case in your state (I speak from our experience in California), but if you go into the high school classroom you will find that kids are not getting writing instruction. Kids are not being asked to write essays much. We did a study which showed that the average number of writing assignments for a tenth-grade class in composition was one a month, and the average length was one page. These are tenth-grade kids. When we asked the teachers, "Why aren't you asking these kids to write more?" the response was "It takes too long to score essays." We have a measurement or performance issue involved; that is, what counts as adequate feedback and how can you provide it to the kids? In another study we did, data came back in a way that showed us some of the kids didn't do well. When we followed up on that, a sizable percentage of the teachers said they didn't re-teach a topic if the students didn't do well; they just dropped it. That response has implications for methodology and for outcomes. Otherwise you may be going through some mindless activity.

I also understand the weakness of the causal chains on what one does instructionally. What concerns me in this focus on process is that people can so easily get hold of the wrong processes. One of the requirements for a certain program is individualized instruction. This is a nice catchword, and everybody can interpret it differently. Through the guidelines it was interpreted to mean that you had to have an individual progress record on each kid. So schools set up "war rooms" with project and status charts. That's okay for projects, but for individual kids it's sort of scary to see that kind of thing up on the bulletin board – reports on each child by subject matter and performance level.

I made some tongue-in-cheek remarks about naturalistic data techniques, but that's exactly what I suggest we do in our evaluation studies: use mixed models, use highly quantitative approaches, and explore ways to develop verifications or hypotheses. We use both outcomes and processes. I think that's the only thing anyone can do right now. Our real concern is not making the evaluation load too onerous to the data provider. For instance, we've been concentrating very strongly on sampling techniques so that *all* the kids or *all* the teachers are not asked to provide information. I still think there's sufficient agreement on some points to allow us to push on outcomes a little bit, but not exclusively. You're right about that.

Question: Would you comment more fully on the role of policy analysts?

Well, it may turn out that evaluators in the future will be differentiated in terms of the ones who move in the general direction of policy analysis, as opposed to those who are more technology oriented – those who do the actual design collection and preliminary interpretation of information. We know that right now the policy people often cast their questions in ways that are unanswerable by data. That may be deliberate. In the cases where it isn't, I think there have been a few instances where policy analysts and evaluators have gotten together on a problem, like an RFP provision, and tried to work out what an evaluation study should have in it. It isn't clear to me how they relate. I wish I could have a pat answer for you. I think policy analysis is something that educational people haven't noticed much before; they assumed that policy analysis work resided at Harvard, Michigan, and Rand, and that was it. I think that as we understand more about it, I think we'll be able to see some changes.

LEADERSHIP SERIES
IN VOCATIONAL AND CAREER EDUCATION

Barlow, Melvin. ***Implications from the History of Vocational Education***, 1976 (OC 15—$1.00).

Bell, Terrel H. ***The Place of Vocational Education in Higher Education: Implications for Education R&D***, 1976 (DL 4—$1.50).

Bell, Terrel H. and Hoyt, Kenneth B. ***Career Education: The U.S.O.E. Perspective***, 1974 (OC 4—$1.50).

Bottoms, James E. ***Implications of the New Vocational Education Legislation for Program Research and Development***, 1976 (OC 23—$1.75).

Buzzell, Charles H. ***Productivity: Implications for Vocational Education***, 1976 (OC 19—$1.00).

Clark, David L. ***Federal Policy in Educational Research and Development***, 1974 (OC 5—$1.50).

Clark, David L. ***Research and Development Productivity in Educational Organizations***, 1978 (OC 41—$2.20).

Cohen, Wilbur J. ***Needed Federal Policy in Education for Century III***, 1977 (OC 24—$1.90).

Cyphert, Frederick. ***Forces Affecting Program Development in Higher Education: Is Anyone Driving?*** 1975 (OC 11—$1.00).

Day, Sherman. ***Education and Training in the Criminal Justice System: Implications for Vocational Education Research and Development***, 1979 (OC 52—$1.90).

Delacruz, Joseph B. ***Educational Programs for Native Americans: Implications for Vocational Education Research and Development***, 1978 (OC 40—$1.90).

Ellis, John. ***Vocational Education and Federal Priorities***, 1978 (OC 47—$1.90).

Ellis, Mary L. ***Vocational Education: The Future Is Now***, 1978 (OC 37—$1.90).

Evans, Rupert. ***Vocational Education R and D in the Past Decade: Implications for the Future***, 1976 (OC 18—$1.00).

Fallstrom, Charles M. ***The Changing Secondary Education Scene: Implications for Vocational Education Research and Development***, 1976 (OC 22—$1.75).

Gideonse, Hendrik. ***A Model for Educational Research and Development: 1985***, 1978 (OC 44—$2.20).

Ginzberg, Eli. ***Strategies for Education Reform***, 1972-73 (DL 1—$1.00).

Gleazer, Edmund J. ***View on Community and Junior College Education***, 1975 (OC 9—$1.00).

Goldhammer, Keith. ***Extending Career Education Beyond the Schoolhouse Walls***, 1974 (OC 3—$2.00).

Halperin, Samuel. ***Emerging Educational Policy Issues in the Federal City: A Report from Washington***, 1978 (OC 42—$2.20).

Herr, Edwin L. ***Work Focused Guidance for Youth in Transition: Some Implications for Vocational Education Research and Development***, 1978 (OC 43—$2.20).

Hicks, Laurabeth L. ***Programs of Guidance and Counseling Becoming of Age: Implications for Vocational Education R&D***, 1977 (OC 25—$1.75).

Hoyt, Kenneth B. ***Career Education, Vocational Education, and Occupational Education: An Approach to Defining Differences***, 1973-74 (DL 2—$1.25).

Jennings, John F. and Radcliffe, Charles W. ***Commentary on Legislation Affecting Vocational Education Research and Development***, 1977 (OC 27—$1.90).

Kolstoe, Oliver P. ***Implications of Research Findings on Vocational and Career Education for the Mentally Handicapped***, 1977 (OC 33—$1.90).

Kottman, Roy M. ***Building a Constituency for Research and Development***, 1975 (OC 10—$1.00).

Krathwohl, David. ***Improving Educational Research and Development***, 1976 (OC 21—$2.20).

Kreitlow, Burton W. ***Trends in Adult Education with Implications for Vocational Education***, 1976 (OC 13—$1.25).

Kruger, Daniel H. ***Occupational Preparation Programs: Implications for Vocational Education***, 1977 (OC 31—$1.90).

Levitan, Sar A. ***The Unemployment Numbers Is the Message***, 1977 (OC 38—$1.90).

McCage, Ronald D. ***The Development of a Comprehensive State Capacity for Program Improvement***, 1978 (OC 34—$1.75).

McCune, Shirley D. ***The Organized Teaching Profession and R&D***, 1977 (OC 29—$1.90).

Marland, Sidney P. ***Career Education: Retrospect and Prospect***, 1974-75 (DL 3—$2.00).

Martin, Edwin. ***New Directions in Vocational Education for the Handicapped: Implications for Research and Devleopment***, 1978 (OC 35—$1.75).

Moody, Tom. ***Vocational Education, CETA, and Youth Unemployment: Meeting the Needs of Inner City Youth***, 1979 (OC 50—$1.75).

Parnes, Herbert S. ***A Conceptual Framework for Human Resource Policy: Implications for Vocational Education Research and Development***, 1976 (OC 14—$1.00).

Petty, Reginald. ***Trends and Issues in Vocational Education: Implications for Vocational Education Research and Development***, 1978 (OC 46—$1.90).

Pucinski, Roman. ***The Role of State and Local Advisory Councils in Vocational Education***, 1978 (OC 36—$1.90).

Reider, Corinne H. ***Women, Work and Vocational Education***, 1977 (OC 26—$1.90).

Rosen, Howard. ***Recent Manpower Legislation and Programs: Implications for Research and Development***, 1975 (OC 7—$1.00).

Scanlon, Robert C. ***Public Schools for the 80's: Implications for Vocational and Career Education R&D***, 1976 (OC 20—$1.50).

Siffin, William. ***Institutional Building in Technical Assistance: The Design Perspective and Some Implications for Research and Development in Vocational Education***, 1975 (OC 8—$1.00).

Simpson, Elizabeth J. ***The Home as a Learning Center for Vocational Development***, 1976 (OC 16—$1.00).

Sticht, Thomas G. ***Literacy and Vocational Competence***, 1978 (OC 39—$2.80).

Wallace, Bertran F. ***Desegregation and Its Implications for Vocational and Career Education***, 1977 (OC 30—$1.75).

Wills, Joan. ***Youth Unemployment: Implications for Vocational Education R&D***, 1977 (OC 32—$1.75).

Wirtz, Willard R. and Ford, Gerald R. ***Bringing the World of Work and the Institutions of Education Closer Together***, 1977 (OC 28—$1.75).

Wirtz, Willard R. ***Community Education Work Councils***, 1976 (OC 17—$1.00).

ORDERING INFORMATION

All prices include postage and handling. When ordering use series numbers and titles. Orders of $10.00 or less will be accepted on a cash, check, or money order basis only. Purchase orders will be accepted for orders in excess of $10.00. Please make check or money order payable to: **The National Center for Research in Vocational Education**. Mail remittance and/or purchase order to: National Center Publications, The Ohio State University, 1960 Kenny Road, Columbus, OH 43210. (Prices subject to change.)

The Lecture Series at the National Center for Research in Vocational Education was established to provide a forum for discussing current issues confronting educational research and development among distinguished professionals and National Center and Ohio State University staff. Points of view or opinions do not necessarily represent official National Center or Ohio State University position or policy.